

Moving Least-Squares Projective Approximation of Manifolds (MMLS)

Barak Sober David Levin

School of Applied Mathematics, Tel Aviv University, Israel

Friday 9th September, 2016

Abstract

In order to avoid the curse of dimensionality, frequently encountered in Big Data analysis, there was a vast development in the field of linear and non-linear dimension reduction techniques in recent years. These techniques (sometimes referred to as manifold learning) assume that the scattered input data is lying on a lower dimensional manifold, thus the high dimensionality problem can be overcome by learning the lower dimensionality behavior. However, in real life applications, data is often very noisy. In this work, we propose a method to approximate a d -dimensional C^{m+1} smooth submanifold \mathcal{M} residing in \mathbb{R}^n ($d \ll n$) based upon scattered data points (i.e., a data cloud). We assume that the data points are located "near" the noisy lower dimensional manifold and perform a non-linear moving least-squares projection on an approximating manifold. Under some mild assumptions, the resulting approximant is shown to be infinitely smooth and of high approximation order (i.e., $O(h^{m+1})$, where h is the fill distance and m is the degree of the local polynomial approximation). Furthermore, the method presented here assumes no analytic knowledge of the approximated manifold and the approximation algorithm is linear in the large dimension n .

1 Introduction

The digital revolution in which we live, have resulted in vast amounts of high dimensional data. This proliferation of knowledge inspires both the industrial and research communities to explore the underlying patterns of these information-seas. However, navigating through these resources encompasses both computational and statistical difficulties. Whereas the computational challenge is clear when dealing with Big-Data, the statistical issue is a bit more subtle.

Apparently, data lying in very high dimensions is usually sparse - a phenomenon sometimes referred to by the name *the curse of dimensionality*. Explicitly, one million data points, arbitrarily distributed in \mathbb{R}^{100} is too small a data-set for data analysis. Therefore, the effectiveness of pattern recognition tools are somewhat questionable, when dealing with high dimensional data [15, 10, 4]. However, if these million data points are assumed to be situated near a low dimensional manifold, e.g., up to six dimensions, then, in theory, we have enough data points for valuable data analysis.

One way to overcome the aforementioned obstacle, is to learn the underlying manifold of the data-set, prior to applying other analysis. Laid in mathematical terms, suppose we

have scattered data points $\{x_i\}_{i=1}^I \subset \mathbb{R}^n$, we wish to find a projection P of $\{x_i\}_{i=1}^I$ onto a d -dimensional manifold (for $d < n$), such that the projected points $s_i = P(x_i)$ maintain the vital information embodied in the original data. The projected data and the manifold itself can later be used for various tasks such as: embedding in a low dimensional linear space, classification, completion of missing data etc.

Perhaps the most well-known dimension reduction technique, presupposing that the data originates from a linear manifold, is the Principal Component Analysis (PCA)[16]. The PCA solves the problem of finding a projection on a linear sub-space preserving as much as possible of the data's variance. Yet, in case the relationships between the scattered data points are more complicated than that, there is no clear-cut solution. The methods used in dimension reduction can range between [20]: linear or non-linear; have a continuous or discrete model; perform implicit or explicit mappings. Furthermore, the type of criterion each method tries to optimize is completely different. For example: *multidimensional scaling* methods [30], *curvilinear component analysis* [9] and *Isomap* [29] aim at preserving distances (either Euclidean or geodesic, local or global) between the data points; *Kernel PCA* methods aim at linearization of the manifold through using a kernel function in the scalar product [27]; *Self Organizing Maps* (SOM) aims at fitting a d -dimensional grid to the scattered data through minimizing distances to some prototypes [31, 17, 18, 20]; *General Topographic Mapping* fits a grid to the scattered data as well, through maximization of likelihood approximation [5, 20]; *Local Linear Embedding* (LLE) aims at maintaining angles between neighboring points [25, 26]; *Laplacian Eigenmaps* approximate an underlying manifold through eigenfunctions of the Graph Laplacian [3]; *Diffusion maps* use the modeling of diffusion processes and utilize Markov Chain techniques to find representation of meaningful structures [8]; and *Maximum Variance Unfolding* uses semi-definite programming techniques to maximize the variance of non-neighboring points [32].

It is interesting to note that all of the aforementioned dimension reduction techniques aim at finding a global embedding of the data into \mathbb{R}^d in a "nearly" isometric fashion. However, as the original manifold is not necessarily isometric or embeddable into \mathbb{R}^d a comprehensive global representation is not feasible. Nash and Whitney embedding theorems state that an embedding is possible into a higher dimensional space - but it is not clear in what manner [28].

Furthermore, albeit the proliferation of methods performing dimension reduction, very little attention have been aimed at denoising or approximating an underlying manifold from scattered data. This pre-processing step could be crucial, especially when the dimension reduction technique being utilized, relies upon differential operators (e.g., eigenfunctions of the graph Laplacian). For clean samples of a manifold a simplicial reconstruction have been suggested as early as 2002 by Freedman [11]. Another simplicial manifold reconstruction is presented in [7], but the algorithm depends exponentially on the dimension. An elaboration and development of Freedman's method utilizing tangential Delauney complexes is presented in [6]. In the latter, the algorithm is linear in the extrinsic dimension, however, no example is presented in the paper. For the case of noisy samples of a manifold there were works aiming at manifold denoising. A statistical approach relying upon graph-based diffusion process is presented in [14]. Another work dealing with locally linear approximation of the manifold is presented in [12].

In our work we assume that our high dimensional data (in \mathbb{R}^n) lies near (or on) a low dimensional smooth manifold (or manifolds), of a known dimension d , with no boundary. We aim at approximating the manifold, handling noisy data, and understanding the local structure of the manifold. Our approach naturally leads to measuring distances from the manifold and to approximating functions defined over the manifold.

The main tool we use for approximating a C^{m+1} smooth manifold is a non-linear Moving Least-Squares approach, generalizing the surface approximating algorithm presented in [22].

The approximation we derive below, is based upon a local projection procedure which results in a C^∞ smooth d dimensional manifold (Theorem 3.17) of approximation order $O(h^{m+1})$, where h is the fill distance (Theorem 3.19). Furthermore, the suggested implementation for this projection procedure is of the complexity order of $O(n)$ (if we neglect the dependency in the lower dimension d). The general idea behind this projection follows from the definition of a differentiable manifold using coordinate charts, collected in a mathematical atlas. The proposed mechanism, takes this concept to the end, and involves the construction of a different coordinate chart for each point on the manifold.

It is worth noting that throughout the article, we use the term smooth manifold to address a submanifold in \mathbb{R}^n , which is smooth with respect to the smoothness structure of \mathbb{R}^n . Explicitly, if a manifold can be considered locally as a smooth graph of an \mathbb{R}^n -valued function, it is said to be smooth.

In Section 2, we start the presentation by reviewing the method of moving least-squares for multivariate scattered data function approximation [23], and its adaptation to the approximation of surfaces from cloud of points [22]. In Section 3 we present the generalization of the projection method of [22] to the general case of approximating a d -dimensional submanifold in \mathbb{R}^n , and in Section 3.3 we discuss the smoothness properties and the approximation power of this projection procedure. We conclude by several examples in Section 4.

2 Preliminaries

As mentioned above, the Moving Least-Squares (MLS) method was originally designed for the purpose of smoothing and interpolating scattered data, sampled from some multivariate function [23, 19, 24]. Later, it evolved to deal with the more general case of surfaces, which can be viewed as a function locally rather than globally [22, 21]. Accordingly, in this brief overview of the topic we shall follow the rationale of [22] and start by presenting the problem of function approximation, continue with surface approximation and in section 3 we conclude with presenting the MLS projection procedure for a general Riemannian submanifold of \mathbb{R}^n .

We would like to stress upfront that throughout the article $\|\cdot\|$ represents the standard Euclidean norm.

2.1 MLS For Function Approximation

Let $\{x_i\}_{i=1}^I$ be a set of distinct scattered points in \mathbb{R}^d and let $\{f(x_i)\}_{i=1}^I$ be the corresponding sampled values of some function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Then, the moving least-squares approximation of degree m at a point $x \in \mathbb{R}^d$ is defined as p_x where:

$$p_x = \operatorname{argmin}_{p \in \Pi_m^d} \sum_{i=1}^I (p(x_i) - f(x_i))^2 \theta(\|x - x_i\|), \quad (1)$$

where $\theta(s)$ is a non-negative weight function (rapidly decreasing as $s \rightarrow \infty$), and $\|\cdot\|$ is the Euclidean norm and Π_m^d is the space of polynomials of total degree m in \mathbb{R}^d . Notice, that if $\theta(s)$ is of finite support then the approximation is made local, and if $\theta(0) = \infty$ the MLS approximation interpolates the data.

We wish to quote here two previous results regarding the resulting approximation presented in [21]. In section 3 we will prove properties extending these theorems to the general case of a d -dimensional Riemannian manifold residing in \mathbb{R}^n .

Theorem 2.1. Let $\theta(t) \in C^\infty$ such that $\theta(0) = \infty$ (i.e., the scheme is interpolatory) and let the distribution of the data points $\{x_i\}_{i=1}^I$ be such that the problem is well conditioned (i.e., the least-squares matrix is invertible). Then the MLS approximation is a C^∞ function interpolating the data points $\{f(x_i)\}_{i=1}^I$.

The second result, dealing with the approximation order, necessitates the introduction of the following definition:

Definition 1. h - ρ - δ **sets of fill distance h , density $\leq \rho$, and separation $\geq \delta$.** Let Ω be a domain in \mathbb{R}^d , and consider sets of data points in Ω . We say that the set $X = \{x_i\}_{i=1}^I$ is an h - ρ - δ set if:

1. h is the fill distance

$$h = \sup_{x \in \Omega} \min_{x_i \in X} \|x - x_i\|, \quad (2)$$

2.

$$\# \{X \cap \overline{B}(y, qh)\} \leq \rho \cdot q^d, \quad q \geq 1, \quad y \in \mathbb{R}^d. \quad (3)$$

Here $\#Y$ denotes the number of elements in a given set Y , while $\overline{B}(x, r)$ is the closed ball of radius r around x .

3. $\exists \delta > 0$ such that

$$\|x_i - x_j\| \geq h\delta, \quad 1 \leq i < j \leq I \quad (4)$$

Remark 2.2. In the original paper [21], the fill distance h was defined slightly different. However, the two definitions are equivalent.

Theorem 2.3. Let f be a function in $C^{m+1}(\Omega)$ with an h - ρ - δ sample set. Then for fixed ρ and δ , there exists a fixed $q > 0$, independent of h , such that the approximant given by equation (1) is well conditioned for θ with a finite support of size $s = qh$. In addition, the approximant yields the following error bound:

$$\|\tilde{p}(x) - f(x)\|_{\Omega, \infty} < M \cdot h^{m+1} \quad (5)$$

Remark 2.4. Although both Theorem 2.1 and Theorem 2.3 are stated with respect to an interpolatory approximation (i.e., the weight function satisfies $\theta(0) = \infty$), the proofs articulated in [21] are still valid taking any compactly supported non-interpolatory weight function.

Remark 2.5. Notice that the weight function θ in the definition of the MLS for function approximation is applied on the distances in the domain. In what follows, we will apply θ on the distances between points in \mathbb{R}^n as we aim at approximating manifolds rather than functions. In order for us to be able to utilize Theorems 2.1 and 2.3, the distance in the weight function of equation (1) should be $\theta(\|(x, 0) - (x_i, f(x_i))\|)$ instead of $\theta(\|x - f(x_i)\|)$ (see Fig. 1). Nevertheless, the proofs of both theorems as presented in [21] are still valid even if we take the new weights. Moreover, the approximation order remains the same even if the weight function is not compactly supported in case the weight function decays exponentially in the fill distance h (e.g., by taking $\theta(r) := e^{-\frac{r^2}{h^2}}$).

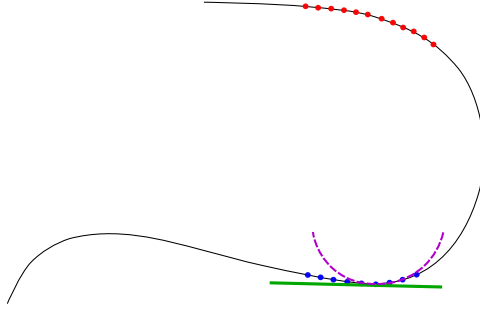


Figure 1: The effect of remote points when taking $\theta(\|(x, 0) - (x_i, f(x_i))\|)$ instead of $\theta(\|x - f(x_i)\|)$. In blue and red we can see samples of a curve. The green line serves as the x -axis for the MLS approximation. The blue points contribute to the cost function $O(h^{m+1})$ whereas the red points contribute $O(\theta(\|x_i - x\|))$. If we define $\theta(r) := e^{-\frac{r^2}{h^2}}$ we get that the red points contribution is $o(h^k)$ for all k . Thus, the red points have a negligible effect over the approximation.

2.2 The MLS Projection For Surface Approximation

Following the rationale presented in [22] let S be an $n - 1$ dimensional submanifold in \mathbb{R}^n (i.e., a surface), and let $\{r_i\}_{i=1}^I$ be points situated near S (e.g., noisy samples of S). Instead of looking for a smoothing manifold, we wish to approximate the projection of points near S onto a surface approximating S . This approximation is done without any prior knowledge or assumptions regarding S , and it is parametrization free.

Given a point r to be projected on S the projection comprises two steps: (a) finding a local approximating n -dimensional hyperplane to serve as the local coordinate system; (b) projection of r using a local MLS approximation of S over the new coordinate system. This procedure is possible since the surface can be viewed locally as a function.

The MLS Projection Procedure

Step 1 - The local approximating hyperplane. Find a hyperplane

$H = \{x | \langle a, x \rangle - D = 0, x \in \mathbb{R}^n\}$, $a \in \mathbb{R}^n$, $\|a\| = 1$, and a point q on H (i.e., $\langle a, q \rangle = D$), such that the following quantity is minimized over all $a \in \mathbb{R}^n$, $\|a\| = 1$, $a = a(q)$:

$$I(q, a) = \sum_{i=1}^I (\langle a, r_i \rangle - D)^2 \theta(\|r_i - q\|) = \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|), \quad (6)$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product in \mathbb{R}^n , and $d(r_i, H)$ is the Euclidean distance between r_i and the hyperplane H . Furthermore, $a(q)$ must be in the direction of the line that passes between q and r , i.e.:

$$(r - q) \parallel a(q). \quad (7)$$

Step 2 - The MLS projection P_m let $\{x_i\}_{i=1}^I$ be the orthogonal projections of the points $\{r_i\}_{i=1}^I$ onto the coordinate system defined by H , so that r is projected to the origin. Referring to H as a local coordinate system we denote the "heights" of the points $\{r_i\}_{i=1}^I$ by $f_i = \langle r_i, a \rangle - D$. We now wish to find a polynomial $p_0 \in \Pi_m^{n-1}$ minimizing the weighted least-squares error:

$$p_0 = \operatorname{argmin}_{p \in \Pi_m^{n-1}} \sum_{i=1}^I (p(x_i) - f_i)^2 \theta(\|r_i - q\|). \quad (8)$$

The projection of r is then defined as

$$P_m(r) \equiv q + p_0(0)a. \quad (9)$$

For an illustration of both Step 1 and Step 2 see Figure 2.

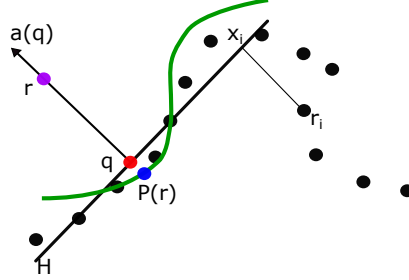


Figure 2: The MLS projection procedure. First, a local reference domain H for the purple point r is generated. The projection of r onto H defines its origin q (the red point). Then, a local polynomial approximation g to the heights f_i of points p_i over H is computed. In both cases, the weight for each of the p_i is a function of the distance to q (the red point). The projection of r onto g (the blue point) is the result of the MLS projection procedure.

As shown in [22] the procedure described above is indeed a projection procedure (i.e., $P_m(P_m(r)) = P_m(r)$). Moreover, let $S \in C^{m+1}$ be the approximated surface and \tilde{S} be the projection's result then, based upon the results obtained in [21] and presented above, it was expected that $\tilde{S} \in C^\infty$ and the approximation order is $O(h^{m+1})$, where h is the mesh size (tending to zero). The approximation order had been proven in [2], however, the C^∞ result was not achieved prior to the current paper. In section 3.3 we present Theorems 3.17 and 3.19 which shows that the approximation is indeed a C^∞ smooth manifold with approximation order of $O(h^{m+1})$ for a more general case.

It is worth mentioning that the most challenging part of the algorithm is finding the approximating hyperplane (i.e., Step 1). The case is so since a depends on q , and the weights are calculated according to the points' distance from q which is a parameter to be optimized as well. It is therefore a non-linear problem. For the full implementation details see [2]. Two examples of surface approximation performed with the MLS projection are presented in Figures 3, 4.

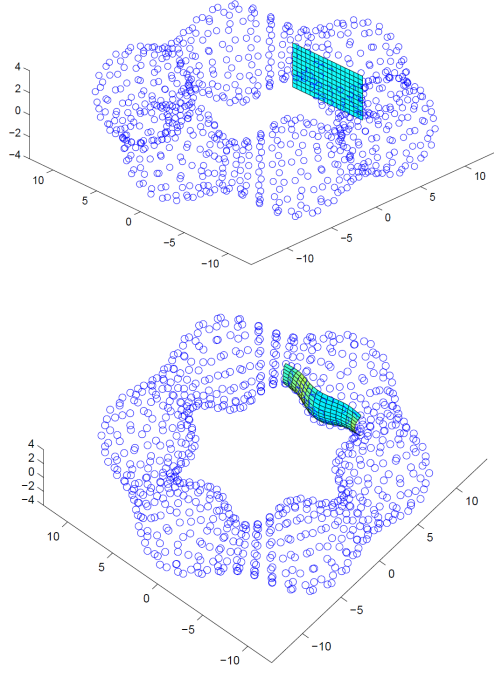


Figure 3: An example of the projection as appeared in [22]: Upper part - data points and a plane segment L near it. Lower part - the projection $P_2(L)$.

3 MLS Projection For Manifolds (MMLS)

The MLS procedure described in the previous section was designed for the case of unorganized scattered points in \mathbb{R}^n lying near a manifold \mathcal{M} of dimension $n - 1$. Here we wish to extend the method to the more general case, where the intrinsic dimension of the manifold is d (for some $d < n$). After presenting the generalized projection algorithm, we propose an implementation, which is linear in the extrinsic dimension n , and conclude with a theoretical discussion.

3.1 The MMLS Projection

Let \mathcal{M} be a C^{m+1} manifold of dimension d lying in \mathbb{R}^n , and let $\{r_i\}_{i=1}^I$ be points situated near \mathcal{M} (i.e., samples of \mathcal{M} with added zero mean noise). We wish to approximate the projection of a point r situated near these samples onto \mathcal{M} (we allow the possibility $r = r_j$ for some $j \in \{1, \dots, I\}$).

Given a point r near \mathcal{M} the projection comprises two steps:

- Find a local d -dimensional affine space $H = H(r)$ approximating the sampled points ($H \simeq \mathbb{R}^d$). We intend to use H as a local coordinate system.
- Define the projection of r using a local polynomial approximation $p : H \rightarrow \mathbb{R}^n$ of \mathcal{M} over the new coordinate system. Explicitly, we denote by x_i the projections of r_i onto H and then define the samples of a function f by $f(x_i) = r_i$. Accordingly, the d -dimensional polynomial p is an approximation of the vector valued function f .

Remark 3.1. Since \mathcal{M} is a differentiable manifold it can be viewed locally as a function from the tangent space to \mathbb{R}^n . It is therefore plausible to assume that we can find a coordinate system H and refer to the manifold \mathcal{M} locally as a function $f : H \rightarrow \mathbb{R}^n$ (see Lemma 3.13 for a formal discussion regarding this matter).

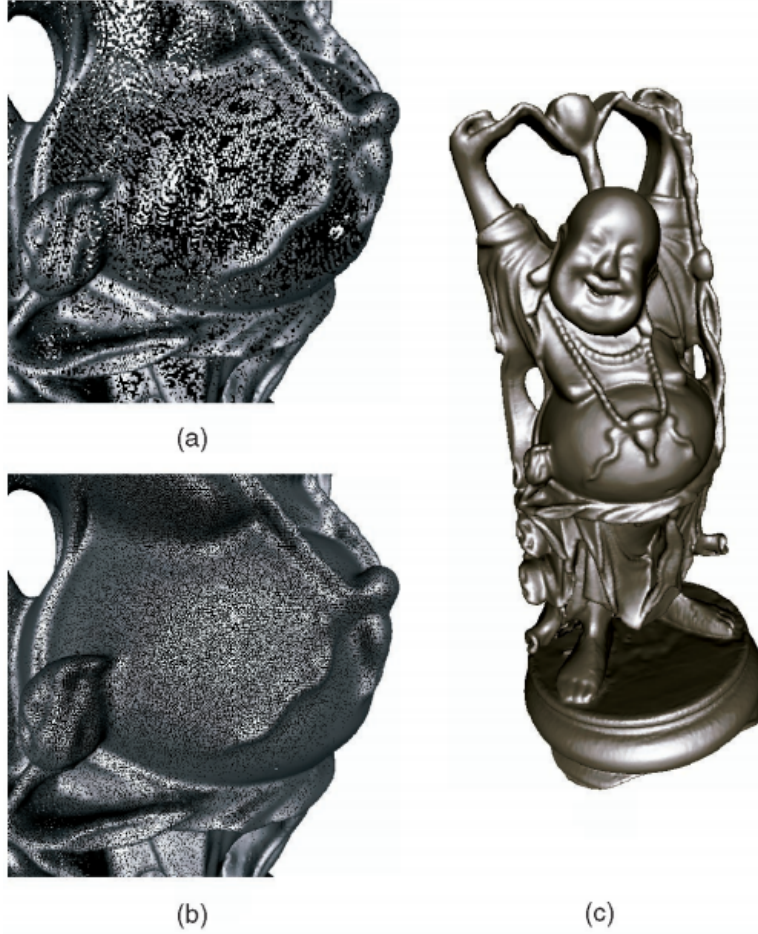


Figure 4: An example of the projection as appeared in [2]: (a) Points acquired by range scanning devices or vertices from a processed mesh typically have uneven sampling density on the surface. (b) Dense and uniform resampling of the object - using the MLS projection . (c) Surface rendering of the object - projecting according to the resampled data.

Remark 3.2. The point r is projected onto a smooth d -dimensional manifold approximating \mathcal{M} . In order to achieve the smoothness property of the approximating manifold H should depend smoothly on r (see Theorem 3.12)

Step 1 - The local Coordinates

Find a d -dimensional affine space H , and a point q on H , such that the following constrained problem is minimized:

$$\begin{aligned}
 J(r; q, H) &= \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|) \\
 \text{s.t.} \quad & r - q \perp H \quad \text{i.e., } r - q \in H^\perp
 \end{aligned} \tag{10}$$

where $d(r_i, H)$ is the Euclidean distance between the point r_i and the subspace H . H^\perp is the $n - d$ dimensional orthogonal complement of H around the origin q , and for the representation of $x \in H$ we use:

$$x = q + \sum_{k=1}^d \alpha_k e_k,$$

where $\{e_k\}_{k=1}^d$ is some basis of the linear space $H - q$.

Assumption 3.3 (Uniqueness domain). *We assume that there exists a subset $U \subset \mathbb{R}^n$ such that*

for any $r \in U$ the minimization problem (10) has a unique solution $q(r) \in U$, and a unique affine subspace $H(r)$, such that the line segment between r and $q(r)$ is in U .

Remark 3.4. For a later use, we introduce the notation $q = q(r)$ and $H = H(r)$ for $r \in U$. Note, that the demand $r - q \perp H$ implies that $q(r)$ would be the same for all r such that $r - q \in H^\perp$ and $r \in U$.

Step 2 - The MLS projection P_m . Let $\{e_k\}_{k=1}^d$ be an orthonormal basis of H (taking q as the origin), and let x_i be the orthogonal projections of r_i onto H (i.e., $x_i = \sum_{k=1}^d \langle r_i, e_k \rangle e_k$ up to a shift by q). As before, we note that r is projected to the origin q . We then denote $f_i = f(x_i) = r_i$, that is we approximate $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$. The approximation of f is performed by a weighted least-squares vector valued polynomial function $\vec{g}(x) = (g_1(x), \dots, g_n(x))^T$ where $g_k(x) \in \Pi_m^d$ is a d -dimensional polynomial of total degree m (for $1 \leq k \leq n$).

$$\vec{g} = \underset{\vec{p} \in \Pi_m^d}{\operatorname{argmin}} \sum_{i=1}^I \|\vec{p}(x_i) - \vec{f}_i\|^2 \theta(\|r_i - q\|). \quad (11)$$

The projection $P_m(r)$ is then defined as:

$$P_m(r) = g(0) \quad (12)$$

Remark 3.5. The weighted least-squares approximation is invariant to the choice of an orthonormal basis of \mathbb{R}^n

Remark 3.6. In fact, considering each coordinate polynomial $g_k(x)$ separately we see that for all $1 \leq k \leq n$ we obtain the same system of least-squares just with different r.h.s. In other words, there is a need to invert (or factorize) the least-squares matrix only once! This fact is important for an efficient application of the implementation for high dimension n .

3.2 Implementation

The implementation of Step 2 is straightforward, as this is a standard weighted least-squares problem. As opposed to that, minimizing (10) is not a trivial task. Since the parameter q appears inside the weight function θ , the weights should be recalculated for each given q . In other words, the problem is non-linear with respect to q . We therefore propose an iterative procedure in which q is updated at each iteration, and the other parameters are solved using a d -dimensional QR algorithm combined with a linear system solver.

Implementation of Step 1 - Finding The Local Coordinates

The main idea here is to perform linear approximations of the data iteratively, and use them as approximations of the affine space H . Assuming we have q_j at the j^{th} iteration, we compute H_{j+1} and then, in view of the goal $r - q \perp H$, we define q_{j+1} as the orthogonal projection of r onto H_{j+1} . However, in order to initiate the process we need a rough first guess. Therefore, we start by taking $q_0 = r$ and solve a spatially weighted PCA around the point r (for more details see (21) in the Appendix). This first approximation is denoted by H_1 and is given by the span of the first d principal components $\{u_k^1\}_{k=1}^d$. Thence, we compute:

$$q_1 = \sum_{k=1}^d \langle r - q_0, u_k^1 \rangle u_k^1 + q_0 = q_0.$$

Upon obtaining q_1, H_1 we can start the following iterative procedure:

- Assuming we have H_j, q_j (the approximation of H and its respective frame $\{u_k^j\}_{k=1}^d$ and origin q_j) we project our data points r_i onto H_j and denote the projections by x_i . Then, we find a linear approximation of the samples $f_i^j = f^j(x_i) = r_i$:

$$\vec{l}^j(x) = \operatorname{argmin}_{\vec{p} \in \Pi_1^d} \sum_{i=1}^I \|\vec{p}(x_i) - f_i^j\|^2 \theta(\|r_i - q_j\|). \quad (13)$$

Notice, that this is a standard weighted linear least-squares as q_j is fixed!

- Given $\vec{l}^j(x)$ we obtain a temporary origin:

$$\tilde{q}_{j+1} = \vec{l}^j(0).$$

Then, around this temporary origin we build a basis $\hat{B} = \{v_k^{j+1}\}_{k=1}^d$ for H_{j+1} in the following manner:

$$v_k^{j+1} := \vec{l}^j(u_k^j) - \tilde{q}_{j+1}$$

We then use the basis \hat{B} in order to create an orthonormal basis $B = \{u_k^{j+1}\}_{k=1}^d$ through a d -dimensional QR decomposition, which costs $O(d^3)$ flops. Finally we derive

$$q_{j+1} = \sum_{k=1}^d \langle r - \tilde{q}_{j+1}, u_k^{j+1} \rangle u_k^{j+1} + \tilde{q}_{j+1}.$$

This way we ensure that $r - q_{j+1} \perp H_{j+1}$.

See Figure 5 for approximated local coordinate systems H obtained by Step 1 on noisy samples of a helix.

Complexity of the MMLS Projection

Since the implementation of Step 2 is straightforward, its complexity is easy to compute. The solution of the weighted least-squares for an m^{th} total degree d -dimensional scalar valued polynomial, involves solving $\binom{m+d}{d}$ linear equations (since this is the dimension of Π_m^d), which is $O(d^m)$ equations for small m . Even though we are solving here for an \mathbb{R}^n -valued polynomial the least-squares matrix is the same for all of the dimensions. Thus, the complexity of this step is merely $O(d^{3m} + n \cdot d^m)$. In addition, we need to compute the distances from the relative origin q^* which costs $O(n \cdot I)$, where I is the number of points. This can be reduced, if we have a compactly supported weight function. Therefore, the overall complexity of the implementation of Step 2 is $O(n \cdot \tilde{I} + d^{3m} + n \cdot d^m)$, where \tilde{I} is the number of points in the support of the weight function.

In a similar way, the complexity of each iteration of Step 1 involves $O(n \cdot \tilde{I} + d^3)$ flops; from our experiments with the algorithm 2-3 iterations are sufficient. However the initial guess of Step 1 involves a PCA which classically costs $O(n \cdot \tilde{I}^2)$. However, as we assume that the data is of intrinsic dimension d we can use a randomized rank d SVD implementation such as the one detailed in [1] and reduce the complexity of this step to $O(n \cdot \tilde{I}) + \tilde{O}(n \cdot d^2)$, where \tilde{O} neglects logarithmic factors of d .

Corollary 3.7. *The overall complexity for the projection of a given point r onto the approximating manifold is $O(n \cdot \tilde{I} + d^{3m} + n \cdot d^m)$. Therefore, the approximation is linear in the large dimension n .*

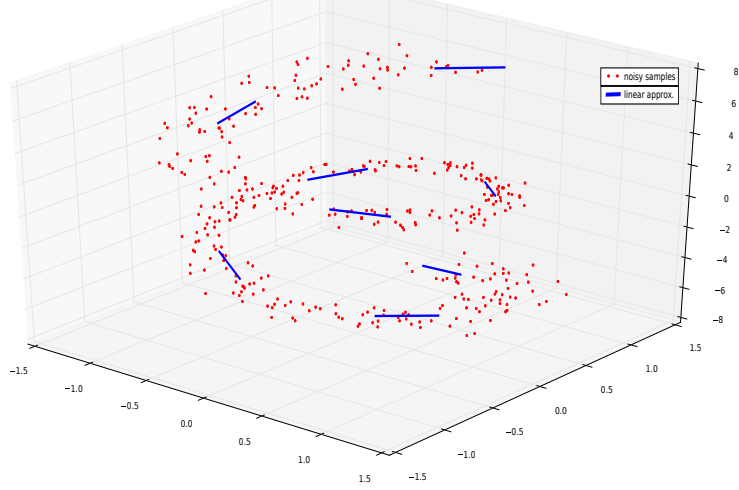


Figure 5: An approximation of the local coordinates resulting from Step 1 implementation after three iterations, performed on several points on a noisy helix.

3.3 Smoothness and Approximation Order of the Approximation

We now wish to define the approximating manifold as $\mathcal{S} = \{P_m(x)|x \in \mathcal{M}\}$, where $P_m(x)$ is the MMLS projection described in equation (12). In this subsection we intend to show that this approximant, is a C^∞ d -dimensional manifold, which approximates the original manifold up to the order of $O(h^{m+1})$, in case of clean samples. Furthermore, we show that $P_m(r) \in \mathcal{S}$ for all r close enough to the sampled manifold \mathcal{M} .

We start our inquiry by showing that the abovementioned procedure is indeed a projection as expected. Following this, we show that the approximating affine spaces $H(r)$ and their origins $q(r)$ are smooth families with respect to the projected point r . As a result we achieve below Theorems 3.17 and 3.19 which gives us the desired properties.

We re-iterate the problem presented in Step 1 and in equation (10): given a point r and scattered data $\{r_i\}_{i=1}^I$, find an affine subspace H of dimension d and an origin $q \in \mathbb{R}^n$ which minimizes

$$J(r; q, H) = \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|),$$

under the constraint

$$r - q \perp H.$$

We denote henceforth the solution to this minimization problem by $q^*(r)$ and $H^*(r)$.

Lemma 3.8. *Let q be fixed. Then H' minimizing the function $J(r; q, H)$ such that $q \in H'$ and $r - q \perp H'$ is determined uniquely by q and can be found analytically. In other words we can write the minimizing H as a function of q , i.e. as $H'(q)$.*

Proof. Let W be the affine 1-dimensional subspace spanned by $r - q$. Specifically, we mean that $W = \text{Span}\{r - q\} + q$. Without loss of generality, we assume $q = \vec{0} \in \mathbb{R}^n$ (otherwise we can always subtract q and the proof remains the same) and therefore H' is now a standard linear space around the origin. Accordingly, since $r - q = r$ the constraint mentioned above can now be rewritten as

$$r \perp H.$$

So now $W = \text{Span}\{r\}$ and we denote the projections of $\{r_i\}_{i=1}^I$ onto W^\perp as $\{w_i\}_{i=1}^I$. Now let $\{e_k\}_{k=1}^n$ be an orthonormal basis of \mathbb{R}^n such that $\{e_k\}_{k=1}^d$ is a basis of H and $e_{d+1} = \frac{r}{\|r\|}$. Using

this notation the minimization problem can be articulated as

$$J(r; q, H) = \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|) = \sum_{i=1}^I \theta(\|r_i\|) \sum_{k=d+1}^n |\langle r_i, e_k \rangle|^2.$$

Looking closer at the inner product on the right hand side we get

$$\sum_{k=d+1}^n |\langle r_i, e_k \rangle|^2 = \|Qr_i - r_i\|^2,$$

where Q is an orthogonal projection of r_i onto H . Now since $H \subset W^\perp$ the first element of this summation

$$|\langle r_i, e_{d+1} \rangle|^2 = \left| \left\langle r_i, \frac{r}{\|r\|} \right\rangle \right|^2,$$

is invariant with respect to the choice of H . Thus we can reformulate the minimization problem as

$$\hat{J}(r; q, H) = \sum_{i=1}^I \theta(\|r_i\|) \sum_{k=d+2}^n |\langle r_i, e_k \rangle|^2 = \sum_{i=1}^I \|Pw_i - w_i\|^2 \theta(\|r_i\|),$$

where P is an orthogonal projection from W^\perp onto H . So in fact we wish to find a projection P^* onto a d -dimensional linear subspace that minimizes the following:

$$\sum_{i=1}^I \|Pw_i - w_i\|^2 \theta(\|r_i\|).$$

From the discussion about the geographically weighted PCA in the Appendix we know that the solution of the original problem is given by taking the span of the first d principal components of the matrix

$$R = \begin{bmatrix} w_1 \cdot \theta(\|r_1\|) & \cdots & w_I \cdot \theta(\|r_I\|) \\ \vdots & & \vdots \end{bmatrix},$$

to be H' - see equation (21). In case $q \neq \vec{0}$ the matrix R will be:

$$R = \begin{bmatrix} \tilde{w}_1 \cdot \theta(\|r_1 - q\|) & \cdots & \tilde{w}_I \cdot \theta(\|r_I - q\|) \\ \vdots & & \vdots \end{bmatrix},$$

where $\{\tilde{w}_i\}_{i=1}^I$ are the projections of $\{r_i - q\}_{i=1}^I$ onto W^\perp . If we denote the singular value decomposition of R by $R = U\Sigma V^T$ and \vec{u}_i are the columns of the matrix U then H' is given explicitly by:

$$H'(q) = \text{Span}\{\vec{u}_i\}_{i=1}^d.$$

□

Remark 3.9. Notice that the rank of the matrix R should be at least d as the dimension of \mathcal{M} is d . In fact, this is an assumption on the local distribution of the data points.

As a result, the minimization problem reduces to minimization with respect to q of

$$J^*(r; q) = J(r; q, H'(q)), \quad (14)$$

where $H'(q)$ can be computed as in Lemma 3.8. This simplifies the minimization task significantly, from the analytic perspective rather than the practical one, as the computation of *SVD* is costly when dealing with large dimensions.

We now wish to tackle the question whether the approximant defined here is indeed a projection operator. In other words, can we say that we project an n dimensional space onto a d dimensional one? In order for this to be true, we must demand that for a sufficiently small neighborhood, elements from H^\perp must be projected onto the same point (see Fig 6 for an illustration). This result is articulated and proved in the following Lemma:

Lemma 3.10. *Let r be in the uniqueness domain of assumption (3.3) U and let $q^*(r)$ and $H^*(r)$ be the minimizers of $J(r; q, H)$ as defined above. Then for any point $r_1 \in U$ s.t. $r_1 - q^*(r) \in H^*(r)^\perp$ we get $q^*(r_1) = q^*(r)$ and $H^*(r_1) \equiv H^*(r)$*

Proof. The result follows immediately from the uniqueness assumption (3.3) and the fact that $r_1 - q^*(r) \perp H^*(r)$ \square

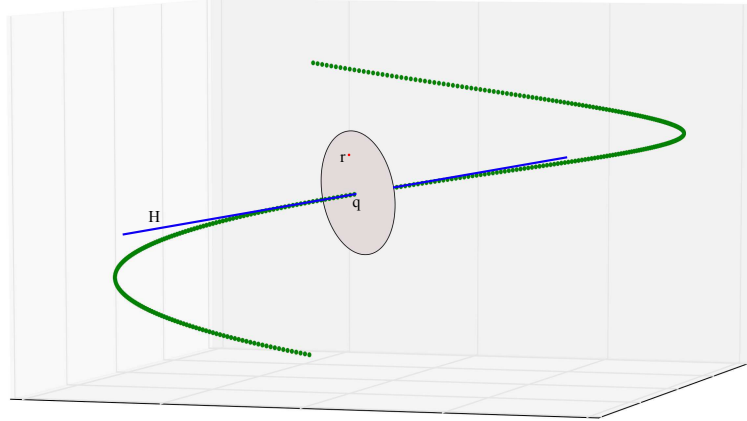


Figure 6: An illustration of a neighborhood of q on H^\perp . All the points in this neighborhood should be projected to the same point.

In order to be able to conduct an in-depth discussion regarding the smoothness of the approximant, and generalize the results quoted in Theorems 2.1 and 2.3, we wish to introduce a definition of a smooth family of affine spaces.

Definition 2. *We say that the affine spaces $H(r)$ **change smoothly** around r if there exist a neighbourhood $A \subset \mathbb{R}^n$ of r , and a set of smooth functions $\Phi(r) = \{\phi_k(r) \in C^\nu(A)\}_{k=1}^d$ such that $\Phi(r)$ is an orthonormal basis of $H(r)$ for all $r \in A$.*

In other words, there is a smooth choice of a moving frame around r .

Lemma 3.11. *Let $\{r_i\}_{i=1}^I$ be noisy samples of a d -dimensional smooth manifold $\mathcal{M} \in C^{m+1}$. Let $\theta(t) \in C^\infty$ be a compactly supported weight function with a support of size $O(h)$, where h is the fill distance. Let the distribution of the data points $\{r_i\}_{i=1}^I$ be such that the minimization problem of $J(r; q, H)$ is well conditioned locally (i.e., the local least-squares matrices are invertible). In addition, let $H'(q)$ be the affine subspace minimizing $J(r; q, H)$ for a given q as described in Lemma 3.8. Then $H'(q)$ changes smoothly (C^∞) with respect to q in a neighbourhood of \mathcal{M} .*

Proof. Let \tilde{q} belong to a neighbourhood of \mathcal{M} , we wish to show that $H'(q)$ is C^∞ smooth in a neighbourhood of \tilde{q} . We know that the original manifold \mathcal{M} , which we are aiming to approximate, is a differentiable manifold. Therefore, locally it can be viewed as a graph of a function from the tangent space $T_p\mathcal{M}$. If we define $p(q)$ to be the projection of q onto \mathcal{M} we get that the samples $\{r_i\}_{i=1}^I$ in the support of \tilde{q} (i.e., with respect to the weight function θ) can be viewed as noisy samples of a function from $T_{p(\tilde{q})}$ to \mathcal{M} . Furthermore, there exists a neighborhood N of \tilde{q} such that for all $q \in N(\tilde{q})$ the samples $\{r_i\}_{i=1}^I$ in the support of q are as well noisy samples of a function $f : T_{p(\tilde{q})}\mathcal{M} \rightarrow \mathbb{R}^n$. Therefore, the minimization problem limited to $q \in N(\tilde{q})$ can be formulated locally as a standard weighted least-squares. Explicitly $H'(q)$ coincides with the following linear approximation:

$$l'(q) = \underset{l \in \Pi_1^d(T_{p(\tilde{q})}\mathcal{M})}{\operatorname{argmin}} \sum_{i=1}^I \|r_i - l(x_i)\|^2 \theta(\|r_i - q\|),$$

where, x_i are the projections of r_i onto $T_{p(\tilde{q})}\mathcal{M}$.

This situation fits the conditions of Theorem 2.1 (in a non-interpolatory setting as discussed in Remark 2.4). Accordingly, we obtain that $l'(q)$, and thus $H'(q)$ as well, belongs to the class C^∞ in the sense of the above mentioned definition. \square

Theorem 3.12. *Let $\theta(x) \in C^\infty$, H be a d -dimensional affine space around an origin q and let $q^*(r)$ and $H^*(r)$ be the minimizers of*

$$J(r; q, H) = \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|)$$

under the restriction

$$r - q \perp H,$$

where $d(r_i, H)$ denotes the Euclidean distance between the point r_i to the affine space H . Then if the matrix $\left(\frac{\partial^2 J^}{\partial q_i \partial q_j}\right)_{ij} \in M_{n \times n}$ is invertible (where $J^*(r; q)$ is the function described in equation (14)) we get:*

1. $q^*(r)$ is a smooth (C^∞) function with respect to r .
2. The affine space $H^*(r)$ changes smoothly (C^∞) with respect to the parameter r .

Proof. First we wish to express the minimization problem under the given constraint using the Lagrange multipliers:

$$J'(r; q, H) = \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|) + \sum_{k=1}^d \lambda_k \langle r - q, e_k(H) \rangle,$$

where $\{e_k(H)\}$ is an orthonormal basis of H . Since we know from Lemma 3.8 that q^* is as well the minimizer of $J^*(r; q)$, which is a function of r and q alone, we can write down:

$$J^*(r; q) = J(r; q, H'(q)) = \sum_{i=1}^I d(r_i, H'(q))^2 \theta(\|r_i - q\|) + \sum_{k=1}^d \lambda_k \langle r - q, e_k(H'(q)) \rangle,$$

where $H'(q)$ is the affine space defined in Lemma 3.8. Specifically, we know that $r - q \perp H'(q)$, therefore, q^* is the minimizer of:

$$J^*(r; q) = \sum_{i=1}^I d(r_i, H'(q))^2 \theta(\|r_i - q\|) \quad (15)$$

In addition, by Lemma 3.11 we know that $H'(q)$ changes smoothly (C^∞) with respect to q . Therefore, there exists a C^∞ choice of bases denoted by $\{e'_k(q)\}_{k=1}^d \subset H'(q)$. Accordingly, we get that $d(r_i, H'(q)) \in C^\infty$, and thus $J^*(r; q) \in C^\infty$ with respect to r and q . Let us now denote:

$$\nabla_q J^*(r, q) = \left(\frac{\partial J^*}{\partial q_1}(r, q), \frac{\partial J^*}{\partial q_2}(r, q), \dots, \frac{\partial J^*}{\partial q_n}(r, q) \right)^T,$$

where q_i is the i^{th} coordinate of the vector $q \in \mathbb{R}^n$. Stated explicitly $\nabla_q J^*$ is a C^∞ function of $2n$ variables:

$$\nabla_q J^*(r, q) : \mathbb{R}^{2n} \rightarrow \mathbb{R}^n.$$

Since q^* minimizes $J^*(r, q)$ for a given r we get:

$$\nabla_q J^*(r, q^*) = 0.$$

Moreover, we know that the matrix

$$\left(\frac{\partial(\nabla_q J^*)_i}{\partial q_j} \right)_{ij} = \left(\frac{\partial^2 J^*}{\partial q_i \partial q_j} \right)_{ij} \in M_{n \times n}$$

is invertible, thus we can apply the Implicit Function Theorem and express q^* as a smooth function of r , i.e. $q^*(r) \in C^\infty$. Moreover, using Lemma 3.11 it follows that $H^*(r) = H'(q^*(r)) \in C^\infty$ as well. □

After establishing the fact that the coordinate system varies smoothly, we turn to the final phase of this discussion, which is the smoothness and approximation order arguments regarding the approximant, resulting from the two-folded minimization problem presented in equations (10)-(11). Initially we wish to approve the fact that the local coordinate system (found by the solution to the minimization problem) is a valid domain for the polynomial approximation performed in Step 2. Ideally, we would have liked to obtain the tangent space of the original manifold as our local coordinate system. In Lemma 3.13 we show that our choice of coordinate system approximates the tangent space and therefore can be considered as a feasible choice for a local domain. Subsequently, we utilize the results articulated in the preliminaries section (i.e., theorems 2.1 and 2.3) to show that we project the points onto a C^∞ manifold, and that given clean samples of \mathcal{M} , these projections are $O(h^{m+1})$ away from the original manifold \mathcal{M} .

Lemma 3.13. *Let $\theta(x)$ be a fast decaying weight function, $\{r_i\}_{i=1}^I \subset \mathcal{M}$ be clean data points and $r \in U$. Furthermore, we denote the linear approximation $H^*(r), q^*(r)$ resulting from the constrained minimization problem of equation (10) and assume that for any $r \in U$ there is a unique projection, $p(r)$ onto \mathcal{M} . Then H^* approximates the data r_i in a neighborhood \mathcal{N} of q^* in the following sense:*

$$d(r_i, H^*) = \sqrt{\|r_i\|^2 - \left\| \sum_{k=1}^d \langle r_i - q^*, e_k \rangle e_k \right\|^2} = \sqrt{\sum_{k=d+1}^n |\langle r_i - q^*, e_k \rangle|^2} = O(h^2), \quad r_i \in \mathcal{N}, \quad (16)$$

where $\{e_k\}_{k=1}^d$ is an orthonormal basis of H^* ; $\{e_k\}_{k=d+1}^n$ is an orthonormal basis of H^\perp (taking q^* as the origin); and h is the mesh size as defined in equation (2).

Proof. Let us rewrite equations (10) using the Lagrange multipliers

$$J'(r; q, H) = \sum_{i=1}^I d(r_i, H)^2 \theta(\|r_i - q\|) + \sum_{k=1}^d \lambda_k \cdot \langle r - q, e_k \rangle. \quad (17)$$

We denote the solution of this minimization problem as $q^*(r), H^*(r)$. Thus the cost function from equation (17) is evaluated as

$$J'(r; q^*, H^*) = \sum_{i=1}^I d(r_i, H^*)^2 \theta(\|r_i - q^*\|).$$

Assuming we have a tangent space $T_{p(r)}\mathcal{M}$ at the projection of r onto \mathcal{M} then $r - p(r) \perp T_{p(r)}\mathcal{M}$ and

$$J'(r; p(r), T_{p(r)}\mathcal{M}) = \sum_{i=1}^I d(r_i, T_{p(r)}\mathcal{M})^2 \theta(\|r_i - p(r)\|).$$

Since $T_{p(r)}\mathcal{M}$ is a the tangent linear approximation of the manifold \mathcal{M} , we know that:

$$d(r_i, T_{p(r)}\mathcal{M}) = O(h^2) \quad , \quad \forall r_i \in \mathcal{N}(p(r), h),$$

where $\mathcal{N}(p(r), h)$ is some small neighborhood of \tilde{q} . Now since θ is of finite support we achieve

$$J'(r; p(r), T_{p(r)}\mathcal{M}) = O(h^4).$$

From the fact that H^*, q^* are the minimizers of equation (17) we get

$$J'(r; q^*, H^*) \leq J'(r; p(r), T_{p(r)}\mathcal{M}).$$

Thus,

$$\sum_{i=1}^I d(r_i, H^*)^2 \theta(\|r_i - q^*\|) = J'(r; q^*, H^*) = O(h^4)$$

and there exists a small enough neighborhood $\tilde{\mathcal{N}}(q^*, h)$ such that $\theta(\|x - q^*\|) > \text{const}$ for $x \in \tilde{\mathcal{N}}$. Therefore in the neighborhood $\mathcal{N}(q^*, h) = \tilde{\mathcal{N}}(q^*, h) \cap \mathcal{N}(p(r), h)$ we achieve the approximation order of:

$$\forall r_i \in \mathcal{N}(q^*, h). \quad d(r_i, H^*) = O(h^2).$$

□

Prior to asserting the theorems which deal with approximation order and smoothness, we wish to remind the reader that the approximating manifold is defined as

$$\mathcal{S} = \{P_{\tilde{m}}(x) | x \in \mathcal{M}\}, \quad (18)$$

where $P_{\tilde{m}}(x)$ is the \tilde{m}^{th} degree moving least-squares projection described in equations (10)-(11). The following discussion will result in proving that \mathcal{S} is indeed a d -dimensional manifold, which is C^∞ smooth and approximates the sampled manifold \mathcal{M} . We prove that \mathcal{S} is a d -dimensional manifold by showing that $P_{\tilde{m}} : \mathcal{M} \rightarrow \mathcal{S}$ is diffeomorphism.

Lemma 3.14. *Let $\theta(x)$ be a compactly supported weight function with a support of $O(h)$, where h is the fill distance. Let $\{r_i\}_{i=1}^I$ be noisy samples of a boundaryless d -dimensional manifold \mathcal{M} , let the data points be distributed such that the minimization problem is well conditioned locally and the fill distance h is small enough. Let $r \in U$ and $\mathcal{M} \subset U$, where U is the uniqueness domain of Assumption 3.3, and let H^*, q^* be the minimizers of equation (10). Then, $q^* : \mathcal{M} \rightarrow \mathbb{R}^n$ is injective.*

Proof. If we assume that the Lemma is false then there exists $r_1, r_2 \in \mathcal{M}$ s.t $r_1 \neq r_2$ and $q^*(r_1) = q^*(r_2) = q$. Since H^* is a function of q^* (see Lemma 3.8) we have that $H^*(r_1) = H^*(r_2) = H$ as well. Furthermore, $r_1 = r_2 + O(h)$ as $\theta(x)$ is of support size of $O(h)$ and the minimization problem of equation (10) yields the same result. In addition, according to the constraint presented in equation (10) $r_1 - q \perp H$ as well as $r_2 - q \perp H$. However, by Lemma 3.13 H approximates $T_{p(q)}\mathcal{M}$ in case of clean samples. Following the rationale of the Lemma's proof we can show in our case that H approximates $T_{r_1}\mathcal{M}$ and $T_{r_2}\mathcal{M}$ as well. Thus, since \mathcal{M} is locally a graph of a function over $T_{r_1}\mathcal{M}$ and $T_{r_2}\mathcal{M}$, for sufficiently small h it is a graph of a function over H . This graph should contain both (q, r_1) and (q, r_2) , however, since $r_1 \neq r_2$ this leads to a contradiction. \square

It is easy to verify that $P_{\tilde{m}} : q^*(\mathcal{M}) \rightarrow \mathcal{S}$ is an injection and therefore $P_{\tilde{m}} : \mathcal{M} \rightarrow \mathcal{S}$ as well. Furthermore, from the definition of the MMLS approximation and the proof of Theorem 2.1 in [21] we can deduce that $P_{\tilde{m}}$ can be viewed locally as a C^∞ function (see the end of the proof of Lemma 3.11 where a similar argumentation is being utilized).

Corollary 3.15. *Assuming that $P_{\tilde{m}}^{-1} : \mathcal{S} \rightarrow \mathcal{M}$ is C^1 , we get that $P_{\tilde{m}} : \mathcal{M} \rightarrow \mathcal{S}$ is a diffeomorphism.*

Remark 3.16. The demand that $\theta(x)$ should be compactly supported can be relaxed to be a fast decaying weight function. However, working with this condition complicates the argumentation, hence, we preferred clarity over generality.

Upon obtaining these results, we are now prepared to move on to the main result of this article. Namely, in the following theorem we show that \mathcal{S} is an approximating d -dimensional manifold, which is C^∞ smooth.

Theorem 3.17. *Let $\theta(t) \in C^\infty$ be a fast decaying weight function, let $\{r_i\}_{i=1}^I$ be noisy samples of a boundaryless d -dimensional submanifold \mathcal{M} of \mathbb{R}^n belonging to the class C^{m+1} , and let the data points be distributed such that the minimization problem is well conditioned locally. Let $\mathcal{M} \subset U$, where U is the uniqueness domain of Assumption 3.3. And let $P_{\tilde{m}}^{-1} : \mathcal{S} \rightarrow \mathcal{M}$ belong to the class C^1 . Then the MMLS procedure of degree \tilde{m} described in equations (10)-(11) projects any $r \in U$ onto a d -dimensional submanifold \mathcal{S} of \mathbb{R}^n . Furthermore, \mathcal{S} is C^∞ smooth.*

Proof. The proof comprises the following arguments:

1. \mathcal{S} is a d -dimensional submanifold in \mathbb{R}^n .
2. $\mathcal{S} \in C^\infty$.
3. $\forall r \in U$, $P_{\tilde{m}}(r) \in \mathcal{S}$.

From Corollary 3.15 combined with Remark 3.16 we get that \mathcal{S} is indeed a d -dimensional submanifold in \mathbb{R}^n as a diffeomorphic image of \mathcal{M} . Furthermore, from the same corollary we can deduce that it can be viewed locally as a C^∞ smooth graph of a function. Thus, \mathcal{S} is a C^∞ smooth manifold with respect to \mathbb{R}^n smoothness structure. Lastly, the fact that $\forall r \in U$, $P_{\tilde{m}}(r) \in \mathcal{S}$ follows from the uniqueness domain assumption. \square

Remark 3.18. The demand that \mathcal{M} is in the uniqueness domain U is in fact a condition about the scale of the noise. If the noise proportion is not too large this assumption is met.

Lastly, we show that given clean samples of \mathcal{M} we achieve that \mathcal{S} approximates \mathcal{M} up to the order of $O(h^{\tilde{m}+1})$.

Theorem 3.19. *Let $\{r_i\}_{i=1}^I$ be an h - ρ - δ set sampled from a C^{m+1} d -dimensional submanifold \mathcal{M} . Then for fixed ρ and δ , there exists a fixed $q > 0$, independent of h , such that the MMLS approximation is well conditioned for θ with a finite support of size $s = qh$. In addition, the approximation yields the following error bound locally:*

$$\|S - \mathcal{M}\|_{\text{Hausdorff}} < M \cdot h^{\tilde{m}+1}, \quad (19)$$

where

$$\|S - \mathcal{M}\|_{\text{Hausdorff}} = \max\{\max_{s \in S} d(s, \mathcal{M}), \max_{x \in \mathcal{M}} d(x, S)\}$$

and $d(p, \mathcal{N})$, is the Euclidean distance between a point p and a manifold \mathcal{N} .

Proof. Let $r \in \mathcal{M}$ then the projection $P_{\tilde{m}}(r) \in \mathcal{S}$ is $O(h^{\tilde{m}+1})$ away from the manifold \mathcal{M} in a neighborhood of r as a local polynomial approximation of a function (see Theorem 2.3). Accordingly, for all $r \in \mathcal{M}$, $d(r, \mathcal{S}) \leq O(h^{\tilde{m}+1})$. Furthermore, for each $s \in \mathcal{S}$ there exists a point $r \in \mathcal{M}$ such that $s = P_{\tilde{m}}(r)$ which is $O(h^{\tilde{m}+1})$ away from \mathcal{M} . Thus, for all $s \in \mathcal{S}$, $d(s, \mathcal{M}) \leq O(h^{\tilde{m}+1})$ as well, and the theorem follows. \square

Remark 3.20. Although entire Section 3 was pronounced using the standard Euclidean norm, all of the definitions, development and proofs are applicable for the general case of an inner product norm of the form $\|\cdot\|_A = \sqrt{x^T A x}$. Where A is a symmetric positive definite matrix. This sort of metric is being utilized in the helix example.

4 Examples

In this section we wish to present some numerical examples which demonstrate the validity of our method. In all of the following examples we have implemented Step 1 as described in Section 3.2 using just three iterations. The weight function utilized in all of the examples (and many others omitted for brevity) is $\theta(r) = e^{-\frac{r^2}{\sigma^2}}$, where sigma was approximated automatically using a Monte-Carlo procedure:

1. Choose 100 points from $\{r_i\}_{i=1}^I$ randomly
2. For each point:
 - Calculate the minimal σ such that the least-squares matrix is well conditioned (in fact we chose 10 times more points than needed).
3. Take the maximal σ from the 100 experiments.

1-dimensional Helix Experiment

In this experiment we have sampled 400 equally distributed points on the helix $(\sin(t), \cos(t), t)$ for $t \in [-\pi, \pi]$ (Fig. 7A) with uniformly distributed (between -0.2 and 0.2) additive noise (Fig. 7B). In all of the calculations we have used the Mahalanobis norm, which is of the type $\sqrt{x^T A x}$, instead of the standard Euclidean. Assigning $d = 1$ (i.e., the manifold's dimension), we projected each of the noisy points and the approximation can be seen in Fig. 7C. The comparison between the approximation and the original as presented in Fig. 7D speaks for itself.

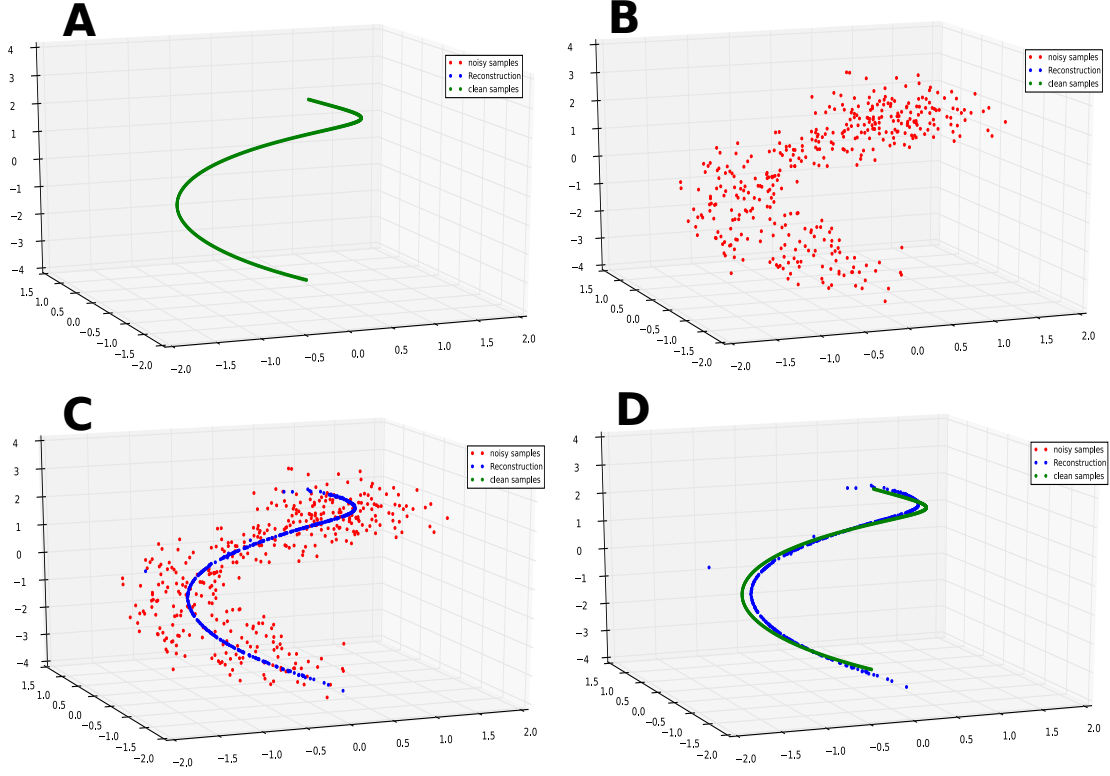


Figure 7: Approximation of 1-dimensional helix. (A) clean samples (green); (B) noisy samples (red), after adding noise distributed $U(-0.2, 0.2)$; (C) the approximation (blue) overlaying the noisy samples (red); (D) comparison between the approximation (blue) and the original clean samples (green)

Ellipses Experiment

Here we sampled 144 images of ellipses of size 100×100 . The ellipses were centered and we did not use any rotations. Thus, we have 144 samples of a 2-dimensional submanifold embedded in \mathbb{R}^{10000} . We have added Gaussian noise $\mathcal{N}(0, 0.05)$ to each pixel in the original images (e.g., see Fig. 8). One of the phenomena apparent n -dimensional data is that if we have a very small random noise (i.e., bounded by ϵ) entered at each dimension, the noise level at the norm level is augmented approximately by a factor of \sqrt{n} . In our case the noise bound is of size $100 \times 0.05 = 5$, whereas the typical distance between neighboring images is approximately $2.5 - 3$. Therefore, if we use the standard Euclidean norm the localization is hampered. In order to overcome this obstacle we have used a 100 dimensional distance. Explicitly, we have performed a pre-processing randomized SVD and reduced the dimensionality to 50 times the intrinsic dimension. The reduced vectors were used just for the purposes of distances computation in the projection procedure process. Several examples of projections can be seen in Fig. 9. An example of the 2 dimensional mapping of the 144 samples projected onto H is presented in 10.

Appendix A - Geographically Weighted PCA

We wish to present here the concept of *geographically weighted PCA* borrowed from [13], as this concept plays an important role in the some of the Lemmas proven in section 3.3 and even in the algorithm itself.

Given a set of n vectors x_1, \dots, x_I in \mathbb{R}^n , we look for a $\text{Rank}(d)$ projection $P \in \mathbb{R}^{n \times n}$ that



Figure 8: Examples of noisy ellipses. The noise is normally distributed $\mathcal{N}(0, 0.05)$ at each pixel.

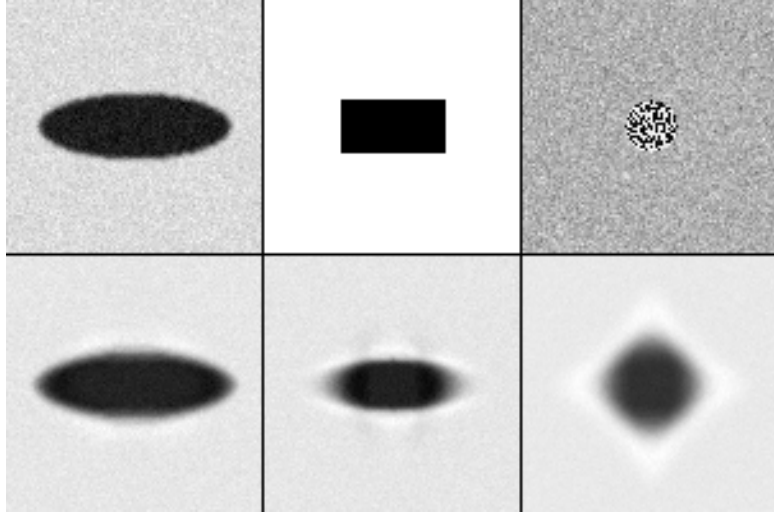


Figure 9: Projections on the ellipses 2-dimensional manifold. Upper line: vectors that were projected. Lower line: the projections of the upper line on the ellipses manifold.

minimizes:

$$\sum_{i=1}^I ||Px_i - x_i||_2^2$$

If we denote by A the matrix whose i 'th column is x_i then this is equivalent to minimizing:

$$||PA - A||_F^2,$$

as the best possible $Rank(d)$ approximation to the matrix A is the SVD $Rank(d)$ truncation denoted by A_d , we have:

$$PA = PU\Sigma V^T = A_d = U\Sigma_d V^T$$

$$P = U\Sigma_d V^T V \Sigma^{-1} U^T$$

$$P = U\Sigma_d \Sigma^{-1} U^T$$

$$P = U I_d U^T$$

$$P = U_d U_d^T$$

And this projection yields:

$$Px = U_d U_d^T x = \sum_{i=1}^d \langle x, u_i \rangle \cdot u_i, \quad (20)$$

which is the orthogonal projection of x onto $span\{u_i\}_{i=1}^d$. Here u_i represents the i^{th} column of the matrix U .

Remark 4.1. The projection P is identically the projection induced by the PCA algorithm.

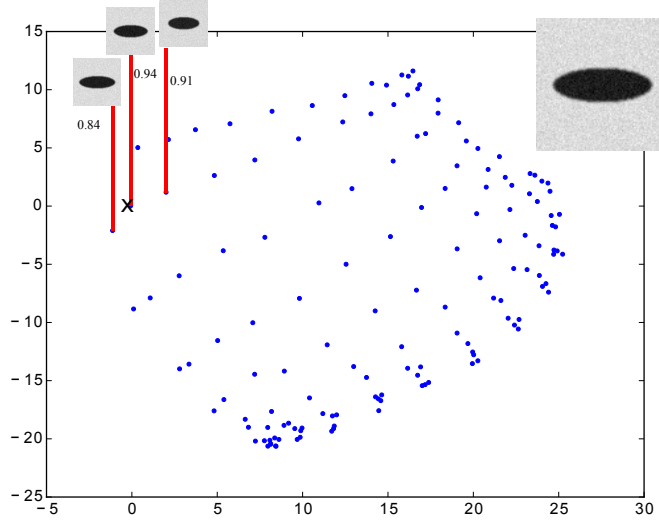


Figure 10: Mapping the ellipses 2-dimensional manifold onto the coordinate system H . In the right upper corner we see the object we wish to project (i.e., r). Marked in \times is the local origin q and some nearby objects from the sampled data alongside their relative weights

The Weighted Projection:

In this case, given a set of n vectors x_1, \dots, x_I in \mathbb{R}^n , we look for a $\text{Rank}(d)$ projection $P \in \mathbb{R}^{n \times n}$ that minimizes:

$$\begin{aligned}
 \sum_{i=1}^I \|Px_i - x_i\|_2^2 \theta(\|x_i - q\|_2) &= \sum_{i=1}^I \|Px_i - x_i\|_2^2 w_i \\
 &= \sum_{i=1}^I \|\sqrt{w_i}Px_i - \sqrt{w_i}x_i\|_2^2 \\
 &= \sum_{i=1}^I \|P\sqrt{w_i}x_i - \sqrt{w_i}x_i\|_2^2 \\
 &= \sum_{i=1}^I \|Py_i - y_i\|_2^2
 \end{aligned}$$

So if we define the matrix \tilde{A} such that the i 'th column of \tilde{A} is the vector $y_i = \sqrt{w_i}x_i$ then we get the projection:

$$P = \tilde{U}_d \tilde{U}_d^T, \quad (21)$$

where \tilde{U}_d is the matrix containing the first d principal components of the matrix \tilde{A} .

References

- [1] Yariv Aizenbud and Amir Averbuch. Matrix decompositions using sub-gaussian random matrices. *arXiv preprint arXiv:1602.03360*, 2016.
- [2] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T Silva. Computing and rendering point set surfaces. *Visualization and Computer Graphics, IEEE Transactions on*, 9(1):3–15, 2003.

- [3] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
- [4] Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, USA, 1 edition, 1957.
- [5] Christopher M Bishop, Markus Svensén, and Christopher KI Williams. Gtm: A principled alternative to the self-organizing map. In *Artificial Neural Networks ICANN 96*, pages 165–170. Springer, 1996.
- [6] Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential delaunay complexes. *Discrete & Computational Geometry*, 51(1):221–267, 2014.
- [7] Siu-Wing Cheng, Tamal K Dey, and Edgar A Ramos. Manifold reconstruction from point samples. In *SODA*, volume 5, pages 1018–1027, 2005.
- [8] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
- [9] Pierre Demartines and Jeanny Hérault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *Neural Networks, IEEE Transactions on*, 8(1):148–154, 1997.
- [10] David L Donoho et al. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pages 1–32, 2000.
- [11] Daniel Freedman. Efficient simplicial reconstructions of manifolds from their samples. *IEEE transactions on pattern analysis and machine intelligence*, 24(10):1349–1357, 2002.
- [12] Dian Gong, Fei Sha, and Gérard Medioni. Locally linear denoising on image manifolds. *Journal of machine learning research: JMLR*, 2010:265, 2010.
- [13] Paul Harris, Chris Brunsdon, and Martin Charlton. Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, 25(10):1717–1736, 2011.
- [14] Matthias Hein and Markus Maier. Manifold denoising. In *Advances in neural information processing systems*, pages 561–568, 2006.
- [15] G Hughes. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14(1):55–63, 1968.
- [16] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [17] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.
- [18] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer Science & Business Media, 2001.
- [19] Peter Lancaster and Kes Salkauskas. Surfaces generated by moving least squares methods. *Mathematics of computation*, 37(155):141–158, 1981.

- [20] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.
- [21] David Levin. The approximation power of moving least-squares. *Mathematics of Computation of the American Mathematical Society*, 67(224):1517–1531, 1998.
- [22] David Levin. Mesh-independent surface interpolation. In *Geometric modeling for scientific visualization*, pages 37–49. Springer, 2004.
- [23] Dermot H McLain. Drawing contours from arbitrary data points. *The Computer Journal*, 17(4):318–324, 1974.
- [24] Andrew Nealen. An as-short-as-possible introduction to the least squares, weighted least squares and moving least squares methods for scattered data approximation and interpolation. URL: <http://www.nealen.com/projects>, 130:150, 2004.
- [25] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [26] Lawrence K Saul and Sam T Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155, 2003.
- [27] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.
- [28] Arkadiy B Skopenkov. Embedding and knotting of manifolds in euclidean spaces. *London Mathematical Society Lecture Note Series*, 347:248, 2008.
- [29] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [30] Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.
- [31] Chr Von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100, 1973.
- [32] Kilian Q Weinberger and Lawrence K Saul. An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In *AAAI*, volume 6, pages 1683–1686, 2006.